DOCUMENT RESUME

ED 418 995 TM 028 287

AUTHOR Hiller, Jack H.
TITLE Applying Computerized Text Measurement Strategies from Project Essay Grade (PEG) to Military and Civilian Organizational Needs.
PUB DATE 1998-04-00
NOTE 8p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Diego, CA, April 13-17, 1998).
PUB TYPE Guides - Non-Classroom (055) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Computer Assisted Testing; Costs; *Human Resources; *Measurement Techniques; Military Personnel; Writing (Composition); *Writing Evaluation
IDENTIFIERS *Project Essay Grade

ABSTRACT

Project Essay Grade (PEG), developed by E. Page (1968) uses computerized methodology to grade student writing. PEG can also be used to assess the quality of civilian and military writing products, such as regulations or instructional texts. Application of PEG offers the potential for enhancing government efforts to monitor writing quality and support human resource management functions. The conduct of writing itself may benefit from a variation of the PEG approach from performance feedback. PEG can be used to monitor various aspects of style, including: (1) eliminating bad construction; (2) avoiding compound constructions; (3) avoiding word-wasting idioms; (4) minimizing use of nominalizations; (5) encouraging the active voice; (6) emphasizing short sentences; (7) using familiar and concrete words; (8) being specific; (9) avoiding exaggeration; and (10) avoiding vagueness. PEG can also be used to improve writing content. Since PEG can be used in a Windows environment and can receive and transmit through the Internet, infrastructure, personnel, maintenance, and update costs should be minimal. (Contains nine references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 418 995

Applying Computerized Text Measurement Strategies from Project Essay Grade (PEG) to Military and Civilian Organizational Needs

Jack H. Hiller, Ph.D., J.D.

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
Jack Hiller
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

TM028287

This paper is prepared for the:
Annual Meeting of the American Educational Research Association in San Diego, CA
April 1998

BEST COPY AVAILABLE

April 1998

Invited paper for Symposium, annual meeting of Amer. Educ. Res. Assoc. San Diego, CA

Applying Computerized Text Measurement Strategies from Project Essay Grade (PEG) to Military and Civilian Organizational Needs

Jack H. Hiller, Ph.D., J.D.
Director of Personnel Technologies, HQ Dept. Army

Ellis Page, having spent a few years as a high school English teacher, learned first hand the time- consuming work required to grade essays for individual classes, and recognized this to be a problem for all English writing teachers and testing agencies (e.g., the College Boards). As an expert in the use of multiple linear regression analysis (Mult-R), Page had wondered if he might apply computer technology to perform the work of measuring various textual features theoretically related to text quality (Proxes) and then apply Mult-R to assign grades. The record shows that the methodology proposed by Page was an instant success (Page and Paulus, 1968), and that subsequent research has continued the initial success (e.g., Page, and Petersen, 1995).

The quality of writings crafted to convey denotative or affective information demonstrates enormous variability across developmental ages and even within large mature professional cohorts. The voluntary popular purchases of the works of select novelists, and even a few textbook writers, attest to the high levels of quality in these writings. However, human grading efforts to assess writing quality, much less precisely measure it, are notoriously burdensome, typically unreliable, and therefore prone to uncertain and low levels of validity. Thus, Page visualized an application of computers to the grading task to save human effort and time while increasing reliability and validity. In addition to using Page's computerized PEG (Project Essay Grade) methodology for grading student writing, it is hard to find any theoretical or practical barriers to using PEG to assess the quality of civilian and military writing products, e.g., regulations, doctrine, instructional texts, etc.. Government writings have for many years been routinely subjected to grading with a variety of readability measures to catch text that needs to be revised. Application of PEG offers the potential for enhancing governmental efforts to monitor writing quality, so poor writing that needs to be re-worked can be intercepted.

MANAGING HUMAN RESOURCES

Valid grading of work product may support important human resource management functions, for example, **selection** for writing jobs, **retention**, selection for **remedial or advanced instruction**, and **promotion**. Since all formal writing is now performed using computers, there is no longer any need to expend time and other resources to convert text to a computer medium. Work product could be routinely monitored, or timely batch processed, to provide personnel managers job performance effectiveness indicators for writing skills.

IMPROVING WRITING EFFECTIVENESS

The conduct of writing instruction may itself benefit from a variation of the PEG approach to computerized text measurement. Most traditional models of instruction for skill learning include *performance feedback* as a critical instructional feature. Writing instruction consists of presentations where principles of effective writing are explained, examples demonstrating effective and ineffective applications of the principles are shown, and then students are given practice exercises to perform, with corrective feedback provided by instructors.

Below, I have illustrated applications of PEG methodology for providing writing feedback on *stylistic quality* and *content accuracy*. The *style* illustrations are largely drawn from Plain English for Lawyers, Richard Wydick (1985) and The Elements of Style, Strunk and White (1965). As will be seen, the instructional feedback for students can also be used by professional writers to spot problems.

STYLE

Eliminate Bad Construction

Wydick distinguishes "working" words from "glue" words. Consider, for example, the following sentence:

> **Sergeant Smith** on **one day tasked** a **Private Jack** to **clean** up **one** of a **number** of **closets** in the **squad** of which **Jack** was a **member.**

Fourteen of twenty seven words, or 52%, are working words, and 13, or 48%, are glue words. Since almost half of the words are glue, we see a problem. By Wydick's scheme, the following words are categorized as glue: **on, a, to, up, of, a, of, in, the, of, which, was** and **a**. Each of these words may be easily identified by computer search, so the computer can provide feedback with reasonable accuracy on the proportion of glue words and the specific glue words it detected. One more balanced version is:

> **SGT Smith tasked PVT Jack** to **clean one** of **Jack's squad's closets.**

Ten of this sentence's twelve words are action words.

Avoid Compound Constructions

In Wydick's terms, compound constructions use three or for words where one or two would suffice. Examples: at that point in time = then; by means of = by; for the purpose of = to. Such discrete compound constructions may be detected with reasonable accuracy by a computer to provide feedback.

Avoid Word-wasting Idioms

Examples: he was aware of the fact that = he knew;
despite the fact that = although;
in many cases = often;
until such time as = until.

Minimize Use of Nominalizations

Action verbs are relatively easy to understand and more effective than their nominalizations. For example, the verb **collide** may be nominalized as **collision.** "The red car collided into the green car," may be nominalized as, "There was a collision between the red and green cars, with the red car initiating the contact between them." The following suffixes provide clues to verb nominalization: al, ence, ancy, ity, ment, ion, ency, ant, ent, and ance.

Prefer the Active Voice to the Passive

This rule is included in Wydick and in Strunk and White. For example, prefer, "**the trial judge ruled**." in place of, "**the ruling was made by the trial judge.**" Passive constructions can be identified by the computer for feedback, and the student can be warned that the computer may sometimes err.

Prefer Short Sentences

Sentence comprehension is positively correlated with brevity. It is interesting to note that sentence length was one of the original Proxes used by Page's Mult-R to predict student writing grades, where it was found to have a significant, positive regression coefficient and a significant, positive bivariate correlation with essay grades assigned by English teachers. Relatively long, grammatically correct sentences indicate general verbal ability and high verbal intelligence (hypotheses verified back in the sixties as a graduate student), and may also directly impress human graders. However, for adult readers, long sentences tend to slow and reduce reading comprehension (Hiller, 1974).

Use Familiar, Concrete Words

Psychological research conducted as far back as the 1960s demonstrated that words rated as concrete *vs* abstract were easier to process and recall. However, relatively long and low familiarity words were found to be positive correlates of writing scores for student writers in PEG research. Writers who use long, unfamiliar words generally have high verbal intelligence and such evidence of superior intelligence may directly create a positive impression in human graders. But readers benefit from short, familiar words (familiar words tend to become short according to the law of least effort to increase efficient communication).

Note that we apparently see here that stylistic advice for raising reading comprehension conflicts with findings on essay grading (i.e., use short sentences and short familiar words to

improve reading comprehension, but use long words and sentences to impress human graders). This conflict may be resolved by recognizing that the specific purpose for writing and the writing cohort's age are variables which control any strategy's success, e.g., children writing to impress the grader *vs* professional writers working to communicate effectively.

Be Specific and Use Examples

The writer's adherence to this recommendation from Strunk and White may be checked by having the computer look for clues to this practice. Specifically ("specifically" is a clue), writers generally precede the explicit use of an example by writing expressions such as: "such as," "for example," "an illustration," "to illustrate," "was illustrated," etc. These cues to the use of examples have been found to be positively correlated with grades for writing quality (Hiller, Marcotte, and Martin, 1969), and test scores of students listening to lectures (Hiller, Fisher, and Kaess, 1969)

Avoid Exaggeration

Violations of Strunk and White's admonition to avoid exaggerations may be checked by having the computer search for clues, such as: always, without any doubt, everyone knows, unique, "most unique," biggest, largest, by far the largest, etc.. These cues for exaggeration have been found to be negatively correlated with grades for writing quality (Hiller, Marcotte, and Martin, 1969).

Avoid Vagueness

Strunk and White recommended avoidance of qualifiers. "Rather, very, little, pretty – these are the leeches that infest the pond of prose, sucking the blood of words." Page 59. The author found it useful in his research (Hiller, 1969) to form 11 categories of vagueness, with a total of 368 words and phrases:

Ambiguity: all of this, other things, other people, somewhere…
Negation: not all, not many, not very…
Approximation: about as, almost, pretty much…
Bluffing and Recovery: clearly, a long story short, as you all know, anyway, of course…
Errors: not sure about, made a mistake, made an error, excuse me…
Indeterminate: a bunch, some, few…
Multiplicity: aspects, factors, sorts, kinds…
Possibility: may, might, chances are, could be…
Probability: probably, sometimes, ordinarily, often, frequently, infrequently…
Reservations: apparently, somewhat, seems, seemingly, tends, tendency…
Anaphora: this, that, them, mentioned before, former, preceding…

CONTENTS

Few courses conducted for military or civilian personnel focus on writing skills, but many courses employ writing exercises for course assignments, and essays to test student learning and retention. All of these forms of writing may benefit from PEG analyses and feedback.

Written essay assignments are used didactically to provide the minimum essential cues required to prompt skilled student performances. Objectively scoreable formats (true/false and multiple choice) may suffer from artificial, tricky, or ambiguous differences between the correct answer and the decoys. Students may become justifiably frustrated and may not learn to form essential memory recall skills and the capabilities required to apply knowledge for job success. However, use of a free format written response can create an enormous burden for instructors, whose grading accuracy may also suffer. So here too, the computer may possibly be employed for grading, only grading of contents instead of, or in addition to, writing style.

Free format test writing requirements may be made to vary from filling in a sentence blank with the right word or phrase, or answering a question with a freely recalled word or phrase, to, for example, speeded writing of elaborate compositions (e.g., legal briefs) over multiple hours without the aid of any notes or other resource materials.

Computer evaluations of essay response accuracy for complex problems (e.g., an advocate's legal brief for overruling an appellate court decision, or the merits and weaknesses of arguments constructed to advocate one side of a dispute and then the other side), are beyond the capability of current PEG or artificial intelligence text evaluation tools. But accurate computer evaluation of relatively short written responses, say two or three sentences that must contain specific key terms, has been successfully demonstrated. Computer assigned scores, using answer keys provided by a history professor for eleven short answer essays in his history course, correlated higher than any single grader's scores with the pooled grades from any 7 of 8 graduate student graders; $r = .60$ vs $r = .48$ for 963 essays, (Hiller, Fisher, and Marcotte,1969).

HOW TO HARNESS PEG FOR STUDENT AND PROFESSIONAL WRITERS IN GOVERNMENT AND MILITARY ORGANIZATIONS

Sometimes optimistic efforts to apply new technologies create inaudible mutterings instead of success stories. "The devil is in the details." "It worked without a hitch in the lab." "They wouldn't listen to us and they wouldn't use it the way we intended." At a minimum, users must be helped to understand how the technology supports their purposes, goals, or objectives. The users must be informed on the technology's pros and cons, and its implementation and operational requirements: operator personnel, equipment, maintenance, software updating, system infrastructure, and student expenses.. Since PEG has been adapted for use in a Windows PC environment and can receive and transmit through the Internet, infrastructure costs, personnel, maintenance and update costs should be minimal.

The Army is now preparing to make quantum increases in its use of Distance Learning, to reduce disruption to work and home life, and to reduce the travel costs generated by resident

school instruction. Distance Learning programs conducted over the Internet will provide a natural medium for applying PEG to both the instructional materials and student writing. PEG should be able to provide a cost effective means for supplementing instructor feedback in Distance Learning programs using the Internet with computer status reports and trend analyses for writing style and content accuracy.

References

Hiller, J. H. (1969). Verbal Response indicators of conceptual vagueness. American Educational Research Journal. 151-161.

Hiller, J.H. (1974). Learning from prose text: Effects of readability level, inserted question difficulty, and individual differences. Journal of Educational Psychology, 66, 201-211.

Hiller, J.H. Fisher, G., and Kaess, E. (1969). Computer investigation of verbal characteristics of effective classroom lecturing. American Educational Research Journal, 661-675.

Hiller, J.H., Fisher, G. and Marcotte, D. (1969). Current perspectives on computer aided content analysis. Invited paper presented to the national meeting of the Association of Computing Machinery, San Francisco, CA.

Hiller, J.H., Marcotte, D., and Martin, T. (1969). Opinionation, vagueness, and specificity-distinctions: Essay traits measured by computer. American Educational Research Journal, 271-286.

Page, E.B., and Paulus, D.H. (1968). The analysis of essays by computer. Final report to the U.S. Dept. of H.E.W., Project No. 6-1318, 1968.

Page, E.B., and Petersen, N. S. (1995). The computer moves into essay grading: Updating the ancient test. Phi Delta Kappan, March, 76(7), 561-565.

Strunk, W., and White, E.B. (1965). The elements of style. MacMillan: New York.

Wydick, R. (1985) Plain English for Lawyers. Carolina Academic Press: Durham, NC.